Evaluating Band Selection for Spectral–Spatial Hyperspectral Classification under Label Scarcity

Sarker Safat Mahmud
Gianforte School of Computing
Montana State University
Bozeman, MT, USA
sarkersafatmahmud@montana.edu

Bradley M. Whitaker
Electrical and Computer Engineering
Montana State University
Bozeman, MT, USA
bradley.whitaker1@montana.edu

John W. Sheppard
Gianforte School of Computing
Montana State University
Bozeman, MT, USA
john.sheppard@montana.edu

***Abstract*—Hyperspectral image classification is important for fields such as precision agriculture, environmental monitoring, food safety, and medical diagnostics. However, it remains challenging due to the large number of spectral bands, insufficient labeled data, and differences in spatial patterns. In this study, we examined how spectral band reduction affects lightweight spectral-spatial classification and how the resulting models behave when labeled data is limited. We compared three band-selection methods, including IBRA+GSS, SSEP, and SRPA, using two lightweight neural network models: Hyper3DNetLite and an attention-based 3D CNN. To ensure a controlled comparison, we applied the same training and hyperparameter tuning process across five widely used benchmark datasets and evaluated spectral settings of 5, 20, and 50 selected bands. Hyper3DNetLite with IBRA+GSS achieved the best overall cross-dataset ranking, although no pipeline showed statistical superiority across all settings. With this setup, we further evaluated pseudo-label-based semi-supervised learning (SSL) when labeled data was limited. Pseudo-labeling provided direct gains in selected low- and intermediate-label settings, while in many other cases its performance remained close to the matched supervised baseline. Overall, the results suggest that pseudo-labeling is most useful when the initial supervised model is sufficiently reliable under the selected band-reduced representation.**

***Index Terms*—Band selection, hyperspectral imaging, label scarcity, pseudo-labeling, semi-supervised learning, spectral–spatial classification.**

I. Introduction

Hyperspectral image (HSI) classification requires models that can handle hundreds of spectral bands while preserving useful spectral and spatial information across different scenes. These images can contain hundreds or even thousands of narrow spectral bands, many of which are strongly correlated. Training deep learning models on such high-dimensional data is computationally expensive and memory-intensive. At the same time, acquiring reliable pixel-level labels is expensive and often requires expert knowledge. Therefore, reducing spectral dimensionality while preserving classification accuracy and reducing the dependence on labeled data are both important problems in hyperspectral image classification.

Spectral-spatial convolutional models address part of this challenge by learning information jointly across spectral channels and neighboring pixels, but their computational requirements can increase with the number of input bands [1]. Earlier work by Morales et al. [2] introduced Hyper3DNet, a hybrid 3D-2D convolutional architecture that uses Principal Component Analysis (PCA) to reduce the spectral dimension. Although the model achieved strong classification performance, the use of multiple 3D convolutional layers made training computationally intensive. In their later work, Morales et al. [3] proposed Hyper3DNetLite, a lightweight variant that evaluates several band-reduction strategies, including Inter-Band Redundancy Analysis with Greedy Spectral Selection (IBRA+GSS) and IBRA+PCA. Their results showed that the IBRA+GSS combination provides effective spectral reduction while retaining high classification accuracy. These results make Hyper3DNetLite a useful representative of classifiers designed for strongly reduced spectral inputs. We therefore include it as one lightweight architecture in this study rather than repeating the classical baseline comparisons already reported in the original work.

More recent attention-based methods such as Spatial–Spectral Edge Preservation (SSEP) and Spectral-Redundancy Penalized Attention Ranking (SRPA) use a different strategy, where spectral importance is learned together with spatial information [4]. However, existing band-selection approaches are commonly evaluated using different classifiers, datasets, and training settings, making it difficult to determine whether performance differences come from the selected bands or from the surrounding experimental setup. We therefore compare IBRA+GSS, SSEP, and SRPA under the same preprocessing, model-training, and evaluation procedure. Hyper3DNetLite and the attention-based 3D CNN provide two complementary lightweight architectures: one was designed for reduced spectral inputs, while the other learns spectral importance through attention.

Another important challenge in HSI is the limited availability of labeled hyperspectral samples. Acquiring pixel-level ground truth is labor-intensive and often requires expert knowledge, making semi-supervised learning (SSL) an important research direction. In this work, we use pseudo-labeling as a simple SSL mechanism to examine whether unlabeled samples provide additional benefit when the manually labeled subset is reduced, rather than comparing different SSL algorithms. Based on the supervised comparison, Hyper3DNetLite with IBRA+GSS is then used for this label-scarcity analysis. Accordingly, this study addresses three questions: how differ-

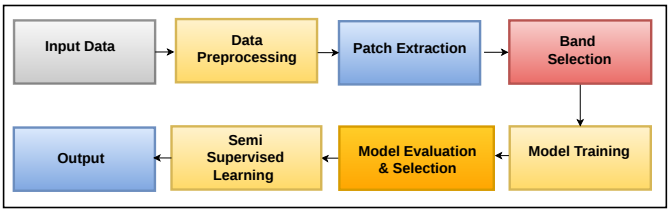


Fig. 1. Overview of hyperspectral classification pipeline.

TABLE I
SUMMARY OF THE HYPERSPECTRAL DATASETS USED IN THIS STUDY.

Dataset	Usable Bands	Classes	Labeled Pixels
Indian Pines	200	16	10,249
Salinas	204	16	54,129
Pavia University	103	9	42,776
Pavia Centre	102	9	148,152
KSC	176	13	5,211

ent band-selection strategies affect lightweight spectral-spatial classifiers under the same experimental conditions; whether any model–band-selection combination remains consistently strong across datasets and spectral budgets; and under what label-scarcity conditions pseudo-labeling provides additional benefit beyond a supervised model trained with the same labeled fraction.

II. Experimental Framework

Figure 1 summarizes the overall experimental workflow.

A. Input Data

We evaluated the methods on five commonly used hyperspectral benchmark datasets [5]: Indian Pines, Salinas, Pavia University, Pavia Centre, and Kennedy Space Center (KSC). Collectively, these datasets represent agricultural, urban, and wetland scenes and cover the visible (VIS), near-infrared (NIR), and short-wave infrared (SWIR) spectral regions. They differ in scene type, spatial resolution, spectral characteristics, class distribution, and degree of class imbalance, allowing the same pipeline to be evaluated under diverse hyperspectral conditions. Table I summarizes the main dataset characteristics.

Indian Pines was acquired using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over an agricultural region in Indiana [6]. The scene contains 16 land-cover classes and is characterized by relatively small and imbalanced classes. Salinas was also acquired using AVIRIS and mainly contains agricultural crops with comparatively larger and more homogeneous regions [7]. Pavia University and Pavia Centre were collected using the ROSIS imaging spectrometer over urban areas in Pavia, Italy [8]. These scenes contain materials such as vegetation, asphalt, buildings, and bare surfaces, with Pavia Centre covering a larger spatial area. KSC was acquired over the Kennedy Space Center wetland environment in Florida and contains vegetation classes with relatively similar spectral characteristics [9]. Together, these datasets provide agricultural, urban, and wetland classification problems under different spectral and spatial conditions. The label distributions are also substantially different: Indian Pines contains several very small classes, including classes with fewer than 50 labeled pixels [6], while KSC contains only 5,211 labeled pixels across 13 spectrally similar wetland classes [9]. These differences become important when only a small fraction of the available labels is retained.

B. Data Preprocessing and Patch Extraction

We preprocessed all hyperspectral datasets using the same dataset-level steps before patch extraction and model training. Each spectral band was independently normalized to the range $[0, 1]$, and background pixels (class 0) were excluded from classification. No additional spatial filtering or denoising was applied so that all band-selection and classification methods used the same spectral information.

Following the spectral-spatial formulation of Morales et al. [2], spatial neighborhoods were extracted around labeled pixels and used as classification samples, with mirror padding applied at image boundaries. Patch size was selected during hyperparameter optimization. Within each cross-validation fold, class-wise sampling was applied only to the training portion, with a maximum of 2000 samples retained per class, while the validation portion preserved its original distribution.

C. Band Selection and Classification Models

We compared three band-selection methods representing different strategies for identifying useful spectral information: IBRA+GSS, SSEP, and SRPA. IBRA+GSS first reduces highly correlated spectral bands using Inter-Band Redundancy Analysis and then applies Greedy Spectral Selection to refine the reduced set [3]. SSEP ranks bands according to the agreement between spatial edges in individual bands and class-boundary information, while SRPA combines attention-derived band importance with a penalty for spectral redundancy [4].

Two lightweight 3D CNN architectures were evaluated: Hyper3DNetLite [3] and the attention-based 3D CNN associated with SSEP/SRPA [4]. Hyper3DNetLite is a reduced-complexity version of Hyper3DNet designed for classification after spectral reduction, while the attention-based model uses a shallow 3D convolutional architecture to learn spectral-spatial representations. Both architectures were used without modification so that the comparison focused on their behavior under different band-selection strategies.

For each model–band-selection combination, spectral budgets of 5, 20, and 50 bands were evaluated.

D. Model Training and Supervised Evaluation

Because the two architectures were originally proposed with different training settings, their hyperparameters were optimized under a common procedure rather than directly reusing either original configuration. Bayesian optimization was performed independently for each dataset using Optuna [10] with a Tree-structured Parzen Estimator (TPE) sampler [11]. The search space included Adam and Adadelta optimizers; patch/stride configurations of $5 \times 5/1$, $9 \times 9/1$, $13 \times 13/1$,

$17\times17/1$, $25\times25/1$, $33\times33/2$, and $50\times50/1$; and batch sizes of $\{8, 16, 32, 64, 128\}$. Batch sizes were restricted to $\{8, 16\}$ for 50×50 patches and to at most 32 for 33×33 patches because of GPU memory requirements. Adam learning rate was searched over $[10^{-4}, 10^{-3}]$, while Adadelta ρ and ϵ were searched over $[0.85, 0.99]$ and $[10^{-7}, 10^{-3}]$, respectively. The original 50×50, stride-25 configuration of the attention-based model [4] was also evaluated separately.

Hyperparameter tuning used stratified 2-fold cross-validation, a maximum of 30 epochs, and early stopping with patience 3. Two folds were used to keep the computational cost of the large search space manageable, while stratification preserved the class distribution. Mean macro F1-score across folds was used as the optimization objective because several datasets are class-imbalanced. The selected architecture–dataset hyperparameters were then fixed, and no further tuning was performed after band reduction.

Final evaluation used stratified 10-fold cross-validation for the six model–band-selection combinations at the three predefined spectral budgets. Band selection was performed independently within each training fold to prevent information leakage, and the same train/validation folds were used across competing methods. Fold-level macro F1-scores were averaged for the cross-dataset statistical analysis. Experiments were run on the Tempest HPC cluster using NVIDIA A40/A100 GPU nodes.

E. Statistical Comparison and Pipeline Selection

To compare the model–band-selection combinations across datasets, each dataset was treated as a block and each of the six combinations as a competing method. Statistical comparison was performed separately for each spectral budget using the Friedman test. The pipelines were ranked within each dataset according to their mean macro F1-score, and the ranks were then averaged across datasets. The null hypothesis was that the competing methods performed equivalently across datasets. When the Friedman test indicated a significant overall difference, Wilcoxon signed-rank tests were performed between the best-ranked pipeline and the remaining pipelines, with Holm correction for multiple comparisons.

The supervised comparison was also used to select the configuration for the label-scarcity experiments. Selection was based on cross-dataset performance and ranking stability rather than an assumption of statistical superiority over all competing methods. Hyper3DNetLite with IBRA+GSS was therefore carried forward because it remained among the strongest configurations across spectral budgets and obtained the best overall cross-dataset ranking. This selection was based on its overall behavior across the datasets rather than a claim that the pipeline is universally superior.

F. Classification under Limited Labeled Data

The second stage examined how the selected spectral-spatial pipeline behaves when only a fraction of the training samples is manually labeled. Hyper3DNetLite with the 5-band IBRA+GSS representation was used for these experiments. Within each training fold, samples were divided into labeled and unlabeled subsets using label ratios of 10%, 20%, 40%, 60%, and 80%. The validation portion remained fully labeled and was not used during pseudo-label generation.

For each label ratio, a supervised baseline was first trained using only the available labeled subset. This provides a matched reference for determining whether adding unlabeled samples provides additional benefit beyond what can already be learned from the same amount of manually labeled data.

We then applied iterative self-training with pseudo-labels to the selected Hyper3DNetLite pipeline. In this approach, sufficiently confident predictions from the current model are assigned as pseudo-labels to unlabeled samples and reused during training [12], [13]. Pseudo-labeling was selected because it provides a simple way to study the contribution of unlabeled data without introducing an additional SSL-specific architecture, augmentation strategy, or loss function. Therefore, this stage is not intended as a comparison of different SSL algorithms.

After the initial supervised training, the model predicted the unlabeled samples. Predictions with a maximum softmax probability greater than 0.9 were added to the labeled training set as pseudo-labeled samples, and the model was retrained. This procedure was repeated for three iterations. The confidence threshold and number of iterations were kept fixed across all datasets rather than tuned separately.

The final pseudo-labeling result at each label ratio was compared directly with the supervised model trained using the same initial labeled subset. This matched comparison separates the additional contribution of unlabeled samples from the effect of simply changing the amount of manually labeled data. Absolute F1-score was also considered, since a small decrease from an already high F1-score is different from poor performance of both models.

III. Results and Discussion

A. Supervised Classification under Band Reduction

The supervised comparison included six competing pipelines evaluated across five datasets and three spectral budgets, resulting in 90 dataset–budget–pipeline evaluations. Table II reports only the best-performing pipeline for each of the 15 dataset–budget settings. Substantial spectral reduction did not prevent strong classification performance across the five benchmark datasets, but no single band-selection strategy was best in every setting. As summarized in Table II, Hyper3DNetLite appeared in 13 of the 15 best-performing dataset–band settings, while the attention-based 3D CNN provided the best result for Pavia University and Salinas in the 5-band setting.

IBRA+GSS appeared more frequently among the best-performing configurations as the spectral budget increased, while SRPA remained competitive across several datasets and SSEP produced strong results in some of the most reduced settings. Therefore, no single band-selection method was optimal for every dataset. This is reasonable because the amount of spectral redundancy and the useful class-discriminative

TABLE II

BEST-PERFORMING CONFIGURATIONS BASED ON MEAN F1-SCORE. B: NUMBER OF BANDS; L: HYPER3DNETLITE; C: ATTENTION-BASED 3D CNN; IBRA: IBRA+GSS. BOLD ENTRIES INDICATE SETTINGS WHERE THE ATTENTION-BASED 3D CNN PRODUCED THE BEST CONFIGURATION.

Dataset	5B	20B	50B
Indian Pines	L+IBRA (0.9844)	L+SRPA (0.9958)	L+IBRA (0.9967)
KSC	L+SSEP (0.9998)	L+SRPA (0.9940)	L+SRPA (0.9962)
Pavia Centre	L+IBRA (0.9954)	L+IBRA (0.9926)	L+IBRA (0.9961)
Pavia University	**C+SSEP (0.9943)**	L+SRPA (0.9954)	L+IBRA (0.9921)
Salinas	**C+IBRA (0.9835)**	L+IBRA (0.9766)	L+IBRA (0.9876)

TABLE III

FRIEDMAN TEST RESULTS COMPARING SIX MODEL–REDUCTION PIPELINES ACROSS FIVE HYPERSPECTRAL DATASETS. AN ASTERISK (*) INDICATES STATISTICAL SIGNIFICANCE AT $\alpha = 0.05$.

Bands	χ_F^2	p-value
5	4.886	0.430
20	7.743	0.171
50	13.000	0.023*

information differ between agricultural, urban, and wetland scenes.

The differences among the six pipelines were further evaluated using the Friedman test. As shown in Table III, the differences were not statistically significant at 5 bands ($p = 0.430$) or 20 bands ($p = 0.171$), while a significant global difference was found at 50 bands ($p = 0.023$). However, the subsequent Wilcoxon signed-rank comparisons with Holm correction did not identify a significant difference between the highest-ranked pipeline and the remaining alternatives. Therefore, these results do not support a claim that one pipeline is statistically superior to all others.

The average ranks in Table IV provide another view of the cross-dataset behavior. Hyper3DNetLite with IBRA+GSS obtained average ranks of 2.4, 2.6, and 1.4 across the three spectral settings, giving it the best overall average ranking. A possible explanation is that Hyper3DNetLite was designed for classification after spectral reduction, while IBRA+GSS removes highly correlated spectral information before classification. Under aggressive reduction, fewer spectral channels are available for the attention mechanism to re-weight, which may partly explain the greater variation observed for some attention-based combinations. The attention-based model nevertheless remained highly competitive and produced the best configuration for two of the 5-band cases. This remains an architectural interpretation rather than a directly tested mechanism.

Based on the cross-dataset performance and ranking stability, Hyper3DNetLite with IBRA+GSS was selected for the label-scarcity experiments. This was an empirical pipeline-selection decision rather than a claim of statistical superiority over all competing methods.

B. Classification under Label Scarcity

Table V compares the supervised baseline and pseudo-labeling results for Hyper3DNetLite with IBRA+GSS across the five labeled-data ratios. Each table entry reports the supervised and SSL F1-scores using the same initial amount of manually labeled data. Bold SSL values indicate settings where pseudo-labeling improved the matched-label supervised baseline.

TABLE IV

AVERAGE RANK OF MODEL–REDUCTION COMBINATIONS (LOWER IS BETTER; BEST OVERALL IN BOLD). L: HYPER3DNETLITE; C: ATTENTION-BASED 3D CNN.

Method	5 Bands	20 Bands	50 Bands	Overall
L + IBRA+GSS	2.4	2.6	1.4	**2.13**
L + SSEP	2.6	3.2	2.8	2.87
L + SRPA	4.4	2.6	3.0	3.33
C + IBRA+GSS	4.2	3.0	5.0	4.07
C + SSEP	3.6	4.6	4.6	4.27
C + SRPA	3.8	5.0	4.2	4.33

Pseudo-labeling showed two useful behaviors under label scarcity. It produced a direct improvement over the matched-label supervised baseline in 8 of the 25 evaluated settings. These improvements were not uniformly distributed across labeled-data ratios. Pseudo-labeling improved 1 of 5 datasets at 10%, 2 of 5 at 20%, 3 of 5 at 40%, 2 of 5 at 60%, and none at 80%. Thus, pseudo-labeling outperformed the matched supervised baseline in 5 of the 10 dataset–ratio cases at the intermediate 40–60% settings, while no improvement was observed at 80%. However, in many other settings the two approaches remained very close. Across all experiments, 15 of the 25 supervised–SSL comparisons differed by no more than 0.01 F1. Therefore, direct improvement is not the only useful behavior under label scarcity. Maintaining similar performance while incorporating unlabeled samples also shows that self-training can remain stable. However, near-equal performance should also not be interpreted as an additional accuracy benefit from unlabeled data; it instead shows that self-training did not substantially degrade the matched supervised result. This pattern suggests that pseudo-labeling may be most useful in an intermediate regime where the supervised model is reliable enough to generate useful pseudo-labels, but enough unlabeled data remains to provide additional training information. At very low labeled ratios, weak initialization can limit self-training, while at high labeled ratios the supervised baseline leaves less room for improvement.

Clearer gains from pseudo-labeling were observed for Pavia University and Salinas in selected low-label settings. Pavia University improved from 0.9346 to 0.9595 at 20% labeled data, while Salinas improved from 0.8833 to 0.9155 at 10%. At larger labeled fractions, the supervised and SSL results generally became similar, consistent with the diminishing headroom observed as more labeled data became available. Indian Pines and KSC showed a different behavior. Their initial supervised performance was much weaker at the smallest labeled fractions, and pseudo-labeling generally could not recover the lost performance. For Indian Pines, several small and imbalanced classes can leave very few representative samples after the labeled subset is reduced. For KSC, the low-label behavior may similarly reflect difficulty representing

TABLE V
SUPERVISED AND SSL MEAN F1-SCORES UNDER DIFFERENT LABELED-DATA RATIOS. BOLD SSL VALUES INDICATE IMPROVEMENT OVER THE MATCHED-LABEL SUPERVISED BASELINE.

Dataset	10%		20%		40%		60%		80%	
	Supervised	SSL	Supervised	SSL	Supervised	SSL	Supervised	SSL	Supervised	SSL
Indian Pines	.0277	.0172	.0354	.0258	.0234	**.0684**	.5219	.4757	.8558	.8353
KSC	.0168	.0133	.3824	.3415	.9647	.9290	.9718	.9408	.9501	.9435
Pavia Centre	.9588	.9573	.9799	.9758	.9846	**.9892**	.9825	**.9888**	.9879	.9782
Pavia University	.8842	.8810	.9346	**.9595**	.9492	**.9530**	.9574	**.9623**	.9740	.9669
Salinas	.8833	**.9155**	.9478	**.9561**	.9779	.9750	.9854	.9807	.9751	.9645

spectrally similar vegetation classes with a small initial labeled subset. Since pseudo-labeling depends on predictions from this initial model, incorrect or uncertain predictions can also limit the usefulness of the unlabeled samples, which is consistent with earlier observations from self-training approaches for hyperspectral classification [12].

Overall, the results suggest three different label-scarcity regimes. At very low labeled ratios, a weak initial classifier can limit the reliability of pseudo-labels, as observed for Indian Pines and KSC. At intermediate ratios, the initial model is more reliable while a substantial unlabeled pool remains, which is where most of the direct SSL gains were observed. At the highest labeled ratio, the supervised baseline was already strong and pseudo-labeling provided little additional improvement. These results suggest that the usefulness of pseudo-labeling depends not only on the amount of labeled data, but also on whether the initial classifier is sufficiently reliable while useful unlabeled samples remain available.

IV. CONCLUSION

The experiments show that strong hyperspectral classification can be maintained after substantial spectral reduction, but the preferred band-selection strategy depends on the dataset and spectral budget. No pipeline was statistically superior across all evaluated settings. Hyper3DNetLite with IBRA+GSS nevertheless obtained the best overall cross-dataset ranking and was therefore used as the operating point for the label-scarcity analysis. The statistical results support this as a stable empirical choice rather than as a pipeline that is significantly superior to every alternative.

Under limited labeled data, pseudo-labeling depended strongly on the initial supervised model. It provided gains in several low- and intermediate-label settings and often remained close to the supervised baseline in other cases, but could not recover performance when the initial model was very weak. At higher labeled ratios, its additional benefit became smaller.

These results suggest a practical sequence: first obtain a reliable band-reduced supervised model with sufficient representation of the available classes, and then use high-confidence pseudo-labels to make use of the remaining unlabeled samples. Pseudo-labeling should not be expected to compensate for a weak initial classifier. Future work will examine confidence-threshold sensitivity and more robust semi-supervised strategies for highly limited or imbalanced labeled subsets.

ACKNOWLEDGMENT

Computational efforts used the Tempest High Performance Computing System supported by Research Cyberinfrastructure at Montana State University.

REFERENCES

[1] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.

[2] G. Morales, J. W. Sheppard, B. Scherrer, and J. A. Shaw, "Reduced-cost hyperspectral convolutional neural networks," *Journal of Applied Remote Sensing*, vol. 14, no. 3, p. 036519, 2020.

[3] G. Morales, J. W. Sheppard, R. D. Logan, and J. A. Shaw, "Hyperspectral dimensionality reduction based on inter-band redundancy analysis and greedy spectral selection," *Remote Sensing*, vol. 13, no. 18, p. 3649, Sep. 2021, doi: 10.3390/rs13183649.

[4] M. I. Karankot, B. M. Whitaker, X. Zhou, and M. U. Masood, "Attention and edge-aware band selection for efficient hyperspectral classification of burned vegetation," in *2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP)*, Aug. 2025, pp. 1–6, doi: 10.1109/MLSP62443.2025.11204320.

[5] X. Z. Zuo, "Hyperspectral data," IEEE DataPort, 2022, doi: 10.21227/3VNQ-ZG28.

[6] M. F. Baumgardner, L. Biehl, and D. A. Landgrebe, "220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian Pine test site 3," Purdue University Research Repository, 2015, doi: 10.4231/R7RX991C.

[7] W. M. Porter and H. T. Enmark, "A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)," in *Imaging spectroscopy II*, vol. 834. SPIE, 1987, pp. 22–31.

[8] B. Kunkel, F. Blechinger, R. Lutz, R. Doerffer, H. van der Piepen, and M. Schroder, "ROSIS (reflective optics system imaging spectrometer) – a candidate instrument for polar platform missions," in *Proc. SPIE*, vol. 868, Apr. 1988, p. 134, doi: 10.1117/12.943611.

[9] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005, doi: 10.1109/TGRS.2004.842481.

[10] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.

[11] Y. Ozaki, Y. Tanigaki, S. Watanabe, M. Nomura, and M. Onishi, "Multiobjective tree-structured parzen estimator," *J. Artif. Intell. Res.*, vol. 73, pp. 1209–1250, Apr. 2022, doi: 10.1613/jair.1.13188.

[12] Y. Wu, G. Mu, C. Qin, Q. Miao, W. Ma, and X. Zhang, "Semi-supervised hyperspectral image classification via spatial-regulated self-training," *Remote Sensing*, vol. 12, no. 1, p. 159, Jan. 2020, doi: 10.3390/rs12010159.

[13] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. ICML 2013 Workshop Challenges in Representation Learning (WREPL)*, Atlanta, GA, USA, Jun. 2013.